

LARGE DEVIATION UPPER BOUNDS FOR SUMS OF POSITIVELY ASSOCIATED INDICATORS

MATTHIAS LÖWE AND FRANCK VERMET

ABSTRACT. We give exponential upper bounds for $\mathbb{P}(S \leq k)$, in particular $\mathbb{P}(S = 0)$, where S is a sum of indicator random variables that are positively associated. These bounds allow, in particular, a comparison with the independent case. We give examples in which we compare with a famous exponential inequality for sums of correlated indicators, the Janson inequality. Here our bound sometimes proves to be superior to Janson's bound.

1. INTRODUCTION

Consider a family of indicator random variables $(X_i)_{i \in I}$ with

$$\mathbb{P}(X_i = 1) = p_i = 1 - \mathbb{P}(X_i = 0).$$

Let $X = \sum_{i \in I} X_i$ be their sum. Moreover, assume that the random variables are associated (or positively associated), i.e. for all coordinatewise increasing functions

$$f, g : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$$

we have that

$$\text{Cov}(f(X_i, i \in I), g(X_i, i \in I)) \geq 0.$$

Association has first been defined and analyzed in [10]. Associated random variables occur in a variety of situations, e.g. in statistical mechanics in the context of the FKG inequalities, in subgraph count statistics in random graphs or in neural networks. For further application it is worth noting, of course, independent random variables are associated and that, if Y_1, \dots, Y_n are associated, and $f_i : \mathbb{R} \rightarrow \mathbb{R}$ is increasing for each $i = 1, \dots, n$, then also Z_1, \dots, Z_n , with $Z_i = f_i(Y_i)$ are associated (see Theorem 2.1 and (P4) in [10]). Here $Y = (Y_1, \dots, Y_n)$. Actually this fact allows to construct and/or discover positive association in many examples.

Date: December 22, 2014.

2000 Mathematics Subject Classification. Primary: 60F10, Secondary: 60C05.

Key words and phrases. Large deviations, positive association, Poisson approximation, Janson inequality.

Example 1.1 (k -Runs). Let $Y_1, Y_2, \dots, Y_{n+k-1}$ be i.i.d. random variables with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$ and for $k \in \mathbb{N}$ let

$$Z_i = \prod_{j=i}^{i+k-1} Y_j$$

be the indicator for a k -run of 1's starting in i . Then, as the Z_i are increasing functions of the independent and thus associated random variables Y_1, \dots, Y_n , they are associated. k -runs are frequently used e.g. in sequence alignments in mathematical biology. We will be interested in the total number Z of k -runs in a Bernoulli sequence of a given length.

Example 1.2 (Subgraph-counts). Let $(Y_{i,j})_{1 \leq i < j \leq n}$ be the i.i.d. Bernoulli random variables occurring in the Erdős-Rényi random graph $G(n, p)$, i.e. $Y_{i,j} = 1$, if and only if the edge between i and j is present in the realization of the random graph, and this occurs with probability p . Otherwise $Y_{i,j}$ is 0. Now let K_k be an arbitrary fixed complete graph, e.g. a triangle for which $k=3$. The random variables

$$Z_{s_1, s_2, \dots, s_k} = \prod_{1 \leq i < j \leq k} Y_{s_i, s_j}$$

are associated. Here the s_i are pairwise different and $\{s_1, s_2, \dots, s_k\} \subseteq \{1, \dots, n\}$. The Z_{s_1, s_2, \dots, s_k} indicate whether the realized graph contains a K_k at the vertices s_1, \dots, s_k . We will be interested in the number of K_k 's in a realization of $G(n, p)$, thus in the sum of the Z_{s_1, s_2, \dots, s_k} , i.e. $Z = \sum_{1 \leq s_1 < \dots < s_k \leq n} Z_{s_1, \dots, s_k}$.

Example 1.3 (U-statistics). Let Y_1, Y_2, \dots, Y_n be i.i.d. random variables with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$ and for $k \in \mathbb{N}$ and $1 \leq i_1, \dots, i_k \leq n$ pairwise different let

$$Z_{i_1, \dots, i_k} = \prod_{j=1}^k Y_{i_j}$$

Then again the Z_{i_1, \dots, i_k} are associated. We will be interested in the U-statistics $Z = \sum_{1 \leq i_1 < \dots < i_k \leq n} Z_{i_1, \dots, i_k}$. Note that the corresponding quantities in Examples 1.1 and 1.2 are structurally similar. However, there Z are so-called incomplete U-statistics, i.e. not all possible random variables Z_{i_1, \dots, i_k} are actually considered (in Example 1.2, e.g., one requires that the vertices form a clique, rather than just multiplying any collection of k edges). This leads to "more independence" among the summands. We will see in Section 3 that the bound we give works particularly well for random variables with many but weak correlations.

Example 1.4 (Random hypergraphs). The following example has several motivations. On the one hand, it is the most generalization of the famous random graph models $G(N, p)$ and $G(N, M)$ invented by Erdős and Rényi ([9]). Here edges are realized between N points i.i.d. with probability p ($G(N, p)$), or we take one of the graphs with exactly M edges at random with equal probability

$(G(N, M))$. These models are readily generalized to hypergraphs in the following way. Given a set of hyperedges, e.g. all possible complete graphs with k vertices that can be constructed on the vertices $\{1, \dots, N\}$ we realize each of them independently at random with probability p or we realize a hypergraph with exactly M such hyperedges at random with equal probability. Such random hypergraphs have been studied e.g. in [7], [3], [4], or [16].

On the other hand, the example can be considered as a mean-field version of the very recent topic of loop-percolation (see e.g. [23]): We cover a graph with a random ensemble of subgraphs and ask for the properties of the thus emerging graph. Finally, our example is also motivated by the analysis of a certain neural network (see [1], [11]). However, it may be regarded as an interesting combinatorial game with a certain similarity to group testing. Take the complete graph K_N and in each of a finite number of steps delete the edges of a randomly chosen complete subgraph K_k (k fixed), if they are not already deleted. Let $Y_{i,j}$ denote the indicator of the event, that edge $\{i, j\}$ has been deleted after n steps. Then the $(Y_{i,j})$ are associated (and so are the $(Z_{i,j}) = (1 - Y_{i,j})$). We will be interested in the probability that the random K_k 's cover all of K_N , hence that the random hypergraph we realize is connected.

Other examples for association have been found in statistical mechanics, e.g. in models obeying the FKG inequalities ([17]) or recently in the fuzzy fractional Potts model [14] or in the parabolic Anderson model [15].

Due to their frequent occurrence, deviation inequalities for positively associated random variables have been investigated in a number of recent articles. Already Newman and Wright prove an invariance principle [18] and martingale inequalities [19]. Cox and Grimmett show a Central Limit Theorem [8]. Boutsikas and Koutras present a simple upper bound for the distance between the distribution of the sum of n associated random variables X_i and a sum of n independent random variables with the same marginals as X_i [6]. Exponential inequalities (also partially together with a resulting Strong Law of Large Numbers) were derived by Oliviera [20], [25], [26], and Yang and Chen [28]. Finally a Law of the Iterated Logarithm was proved by Xing and Yang [27]. A very readable survey over these and more results can be found in [21].

The purpose of the present note is to give a new exponential bound for the probability that a sum of positively associated indicator variables is particularly small or even 0. These bounds will also allow for a comparison with the case of sums of independent indicators with the same distribution. Such a comparison is of interest, since often when positively associated random variables are applied in the literature, it is tacitly assumed that the probabilities of their sum being 0 is well approximated by the independent case (see e.g [1]). Moreover, by positive

association, we have that

$$\mathbb{P}\left(\sum_i X_i = 0\right) \geq \prod_i \mathbb{P}(X_i = 0)$$

(with the sum and the product being taken over the same set of indices), anyway. Thus we always have an obvious bound in one direction by the independent case. One way to obtain the comparison between the dependent and the independent situation is Janson's inequality ([12], [13]). This can be stated as follows: In the above setting, let I be index set of the associated variables $(X_i)_{i \in I}$

$$\begin{aligned} X &= \sum_{i \in I} X_i, & \lambda &= \mathbb{E}X \\ \Delta &= \frac{1}{2} \sum_{\substack{i \sim j, \\ i \neq j}} \mathbb{E}(X_i X_j) & \bar{\Delta} &= \lambda + 2\Delta, \end{aligned}$$

where we write $i \sim j$ if and only if $\text{Cov}(X_i, X_j) \neq 0$. Then the following bounds hold.

Theorem 1.5. ([12], [13], *also see* [24]) *We have*

$$\mathbb{P}(X = 0) \leq e^{-\lambda + \Delta} \quad (1.1)$$

and

$$\mathbb{P}(X = 0) \leq e^{-\frac{\lambda}{\bar{\Delta}^2}}. \quad (1.2)$$

Moreover,

$$\mathbb{P}(X = 0) \leq \exp\left(\frac{\Delta}{1 - \max_{i \in I} \mathbb{E}X_i}\right) \prod_{i \in I} (1 - \mathbb{E}X_i). \quad (1.3)$$

For the purpose of the present note, we will be most of all interested in comparing our results with (1.3), which can be derived from (1.1), but was first proved by Boppona and Spencer [5].

Another way to compare the dependent and the independent cases is the following result, by Boutsikas and Koutras.

Theorem 1.6. ([6]) *We have*

$$\mathbb{P}[X = 0] \leq \prod_{i=1}^n \mathbb{P}[X_i = 0] + \sum_{i < j} \text{Cov}(X_i, X_j). \quad (1.4)$$

Our main theorem also gives an upper bound for the probability that the sum of positively or negatively associated indicator random variables is zero. It can be stated as follows.

Theorem 1.7. *Let the X_i be positively associated indicators. With the above notation we have for all $t > 0$*

$$\mathbb{P}(X = 0) \leq e^{-t|I|} \left(\prod_{i \in I} \mathbb{E}[e^{t(1-X_i)}] + t^2 e^{t|I|} \sum_{i < j} \text{Cov}(X_i, X_j) \right) \quad (1.5)$$

If the (X_i) are identically distributed with $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ this boils down to

$$\begin{aligned} \mathbb{P}(X = 0) &\leq e^{-t|I|}(p + (1-p)e^t)^{|I|} + t^2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ &= (1-p)^{|I|} \left(1 + e^{-t} \frac{p}{1-p}\right)^{|I|} + t^2 \sum_{i < j} \text{Cov}(X_i, X_j) \end{aligned} \quad (1.6)$$

- Remarks 1.8.** (1) It is obvious that Theorem 1.5, Theorem 1.6 as well as our Theorem 1.7 are tailor-made for the situation of weak correlations, where one is close to the independent situation. If correlations are getting large, then they give useless bounds: The probabilities are bounded from above by numbers larger than 1. An important difference between (1.4) and (1.5) is that we have a factor t^2 in front of the sum of the covariances. As we will see in Section 3, choosing t small, this factor can be decisive when the covariances are too large.
- (2) A quick glance at (1.3) and (1.6) already reveals the major difference between the two bounds: While (1.3) is purely multiplicative, our bound (1.6) also has an additive component. This has a consequence for the applicability of our bounds. Consider e.g. the situation for (1.6): In case that $\mathbb{P}(X = 0)$ is comparable to the independent approximation $(1-p)^{|I|}$ and p is much larger than $1/|I|$, both probabilities will typically go to 0 and they will do so at an exponential rate. To have the additive term $t^2 \sum_{i < j} \text{Cov}(X_i, X_j)$ in (1.6) smaller than the probability we want to approximate, we therefore need to choose t extremely small, which in turn means that the multiplicative correction in $(1 + e^{-t} \frac{p}{1-p})^{|I|}$ is of order $e^{p|I|}$ which typically is only effective, if p is of order $1/|I|$. Hence in such cases, we expect that (1.3) will yield results that are more useful than (1.6).

We will prove Theorem 1.7 in the following section. Section 3 will be devoted to comparing Theorems 1.5 and 1.7 in the examples mentioned above. We will see that our bound sometimes is better than Janson's inequality and also better than the bound given in Theorem 1.6.

2. PROOF OF THEOREM 1.7

As usual the central ingredient of the proof of an exponential inequality as in Theorem 1.7 is an estimate of the moment generating function. Such bound can be found in [22], Lemma 3.1. For the sake of completeness (and also as the original proof is somewhat short), we reprove the statement here.

Lemma 2.1. (see [22], Lemma 3.1) Let X_1, X_2, \dots, X_n be positively associated random variables, such that $X_i \leq \kappa$ for all i and some $\kappa > 0$. Then for any $t > 0$

$$\left| \mathbb{E} \left[e^{t \sum_{i=1}^n X_i} \right] - \prod_{i=1}^n \mathbb{E} \left[e^{t X_i} \right] \right| \leq t^2 e^{nt\kappa} \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

Proof. The key tool is to use Lemma 3 in [17]. This Lemma implies that for associated random variables X and Y with finite variance, and (real or complex valued) functions f and g with bounded derivatives one has

$$|\text{Cov}(f(X), g(Y))| \leq \|f'\| \|g'\| \text{Cov}(X, Y)$$

where $\|\cdot\|$ is the sup-norm. Applying this we obtain due to the boundedness of the X_i

$$|\text{Cov}(e^{tX_1}, e^{tX_2})| \leq t^2 e^{2t\kappa} \text{Cov}(X_1, X_2)$$

which is precisely the assertion for $n = 2$. The result now follows by induction: According to our earlier remark, if X_1, X_2, \dots, X_n are associated random variables, then so are $X_1 + X_2 + \dots, X_{n-1}$ and X_n . Thus, if the assertion has already been proven for $n - 1$ we obtain

$$\begin{aligned} & \left| \mathbb{E} \left[e^{t \sum_{i=1}^n X_i} \right] - \prod_{i=1}^n \mathbb{E} \left[e^{t X_i} \right] \right| \\ & \leq \left| \mathbb{E} \left[e^{t \sum_{i=1}^n X_i} \right] - \mathbb{E} \left[e^{t \sum_{i=1}^{n-1} X_i} \right] \mathbb{E} \left[e^{t X_n} \right] \right| + \left| \mathbb{E} \left[e^{t \sum_{i=1}^{n-1} X_i} \right] \mathbb{E} \left[e^{t X_n} \right] - \prod_{i=1}^n \mathbb{E} \left[e^{t X_i} \right] \right| \\ & \leq t^2 e^{nt\kappa} \sum_{1 \leq i \leq n-1} \text{Cov}(X_i, X_n) + t^2 e^{nt\kappa} \sum_{1 \leq i < j \leq n-1} \text{Cov}(X_i, X_j) \end{aligned} \quad (2.1)$$

$$\leq t^2 e^{nt\kappa} \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j), \quad (2.2)$$

where in (2.1) the first summand is bounded by another application of Lemma 3 in [17] inequality, while the bound on the second summand follows from the induction hypothesis. \square

From this bound the proof of the main theorem follows quickly:

Proof of Theorem 1.7. With the notation of the theorem let $Y_i = 1 - X_i$. Note that together with the (X_i) also the (Y_i) are associated. Using an exponential

Markov inequality and the previous lemma we obtain for all $t > 0$

$$\begin{aligned}
\mathbb{P}(X = 0) &= \mathbb{P}\left(\sum_{i \in I} X_i = 0\right) = \mathbb{P}\left(\sum_{i \in I} Y_i = |I|\right) = \mathbb{P}\left(\sum_{i \in I} Y_i \geq |I|\right) \\
&\leq e^{-t|I|} \mathbb{E}[e^{t \sum_{i \in I} Y_i}] \\
&\leq e^{-t|I|} \left(\prod_{i \in I} \mathbb{E}[e^{tY_i}] + t^2 e^{t|I|} \sum_{i < j} \text{Cov}(Y_i, Y_j) \right) \\
&= e^{-t|I|} \left(\prod_{i \in I} \mathbb{E}[e^{t(1-X_i)}] + t^2 e^{t|I|} \sum_{i < j} \text{Cov}(X_i, X_j) \right)
\end{aligned}$$

which is (1.5).

Computing the expectation in the product for identically distributed random indicators with success probability p we arrive at

$$\begin{aligned}
\mathbb{P}(X = 0) &\leq e^{-t|I|} \left(\prod_{i \in I} (p + (1-p)e^t) + t^2 e^{t|I|} \sum_{i < j} \text{Cov}(X_i, X_j) \right) \\
&= e^{-t|I|} (p + (1-p)e^t)^{|I|} + t^2 \sum_{i < j} \text{Cov}(X_i, X_j) \\
&= (1-p)^{|I|} \left(1 + \frac{p}{1-p} e^{-t} \right)^{|I|} + t^2 \sum_{i < j} \text{Cov}(X_i, X_j)
\end{aligned}$$

which is (1.6). □

3. EXAMPLES

In this section we will discuss the quality of our estimate in the examples given in Section 1. The aim is to see, if and when Theorem 1.7 is better than Theorem 1.5, especially, when (1.6) is more powerful than (1.3), and to gain a general understanding, in which situations it is advisable to apply Theorem 1.7.

We start with

Example 3.1 (k-Runs, Example 1.1 continued). To keep the situation symmetric and simplify computations, we slightly change the setup. So let Y_1, Y_2, \dots, Y_n be i.i.d. random variables with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$ and for $k \in \mathbb{N}$ let

$$Z_i = \prod_{j=i}^{i+k-1} Y_j$$

(where $i + k - 1$ now is to be taken modulo n). Moreover let $Z = \sum_{i=1}^n Z_i$. Computing the central quantity in Janson's inequality we obtain

$$\begin{aligned}\Delta &= \frac{n}{2} \sum_{\substack{1 \sim j, \\ 1 \neq j}} \mathbb{E} Z_1 Z_j = \frac{n}{2} \sum_{j=1}^{k-1} p^{k+j} \\ &= \frac{n}{2} p^{k+1} \frac{1 - p^{k-1}}{1 - p}\end{aligned}$$

Thus (1.3) yields

$$\mathbb{P}(Z = 0) \leq (1 - p^k)^n \exp \left(\frac{np^{k+1}(1 - p^{k-1})}{2(1 - p)(1 - p^k)} \right). \quad (3.1)$$

(Note that more generally the total variation distance between the distribution of Z and a Poisson distribution was bounded in [2], Theorem 8.F.) This bound becomes effective, i.e. we have that $\mathbb{P}(Z = 0)$ is well approximated by $(1 - p^k)^n$, if $p \ll n^{-1/k+1}$, i.e. if $np^{k+1} \rightarrow 0$. On the other hand our Theorem 1.7 yields for any $t > 0$

$$\begin{aligned}\mathbb{P}(Z = 0) &\leq (1 - p^k)^n (1 + e^{-t} \frac{p^k}{1 - p^k})^n + t^2 \sum_{i < j} \text{Cov}(Z_i, Z_j) \\ &\leq (1 - p^k)^n (1 + e^{-t} \frac{p^k}{1 - p^k})^n + t^2 \Delta\end{aligned}$$

Choosing $0 < t \ll \Delta^{-1/2}$ the second summand on the right hand side converges to 0. Then the correction factor $(1 + e^{-t} \frac{p^k}{1 - p^k})^n$ for the first summand on the right hand side is bounded by $e^{n \frac{p^k}{1 - p^k}}$ and particularly converges to 1, if $p \ll n^{-1/k}$, which improves the bound from Janson's inequality. Note however that even for much larger p an exponential bound on $\mathbb{P}(Z = 0)$ holds true. As a matter of fact Barbour et al. in [2], Theorem 8.G give the bound

$$|\mathbb{P}(Z = 0) - \exp(-n(1 - p)p^k)| \leq (2k(1 - p) + 1)p^k \quad (3.2)$$

The recipe seen in the previous example illustrates the general method. In (1.6) one always chooses t (depending on n) so small that the second summand on the right vanishes, and then we obtain a bound of the form $(1 - p)^n e^{n \frac{p}{1 - p}}$, where p is the success probability of the random variables in question. This works, if p is so small, that $(1 - p)^n$ is of constant order. However, even if both $(1 - p)^n$ and $\mathbb{P}(Z = 0)$ tend to 0, we may choose t so small, that the second summand in (1.6) vanishes, but we may not be able to guarantee that then p is so small, that $e^{n \frac{p}{1 - p}}$ converges to 1.

Example 3.2 (Subgraph-counts, Example 1.2 continued). In the situation of Example 1.2 consider the number of triangles in the random graph $G(n, p)$. Thus we consider

$$Z_{s_1, s_2, s_3} = Y_{s_1, s_2} Y_{s_2, s_3} Y_{s_3, s_1}$$

and

$$Z = \sum_{1 \leq s_1 < s_2 < s_3 \leq n} Z_{s_1, s_2, s_3}.$$

We bound the probability that the random graph does not contain any triangle, hence that $Z = 0$. The quantity Δ in (1.3) can this time be computed as

$$\Delta = \frac{1}{2} \binom{n}{3} 3np^5 \sim \frac{1}{4} n^4 p^5,$$

as there are $\binom{n}{3}$ triangles, each of which is correlated with $3n$ other triangles, and in this case $\mathbb{E}Z_{s_1, s_2, s_3} Z_{s_4, s_5, s_6} = p^5$. Thus Janson's inequality gives

$$\mathbb{P}(Z = 0) \leq (1 - p^3)^{\binom{n}{3}} \exp\left(\frac{n^4 p^5}{4(1 - p^3)}\right).$$

If $p \ll n^{-4/5}$, this bound is effective, since the factor is tending to 1.

This time our bound works well only in some range of the parameters and for slightly different reasons than in the previous example. From (1.6) we obtain for any $t > 0$

$$\mathbb{P}(Z = 0) \leq (1 - p^3)^{\binom{n}{3}} \left(1 + e^{-t} \frac{p^3}{1 - p^3}\right)^{\binom{n}{3}} + t^2 n^4 p^5.$$

In the regime, where $n^{-1} \ll p$ we see exactly the problem mentioned at the end of the introduction occurring: The probabilities to be approximated are of order $\exp(-p^3 n^3)$ (at least for p going to 0 with n to infinity) and thus converging to 0 exponentially fast. So even if $p \ll n^{-4/5}$, in which case the second term on the right goes to 0, we need to choose t to go to 0 as well, and we cannot improve the first summand in our bound by choosing t large. If now $n^{-1} \ll p \ll n^{-4/5}$ we may obtain a bound of the form $(1 - p^3)^{\binom{n}{3}} \left(1 + \frac{p^3}{1 - p^3}\right)^{\binom{n}{3}}$ by choosing t extremely small. This is however useless, as the second factor on the right diverges.

So let us assume that $p = \lambda/n$. Choose $t = n^\alpha$ for $0 < \alpha < \frac{1}{2}$. Then (1.6) becomes

$$\begin{aligned} \mathbb{P}(Z = 0) &\leq (1 - p^3)^{\binom{n}{3}} \left(1 + e^{-n^\alpha} \frac{p^3}{1 - p^3}\right)^{\binom{n}{3}} + \frac{\lambda^5}{n^{1-2\alpha}} \\ &= (1 - p^3)^{\binom{n}{3}} \left(1 + e^{-n^\alpha} \frac{p^3}{1 - p^3}\right)^{\binom{n}{3}} + o(1). \end{aligned}$$

The multiplicative correction $\left(1 + e^{-n^\alpha} \frac{p^3}{1 - p^3}\right)^{\binom{n}{3}} \leq \exp(\lambda^3 e^{-n^\alpha})$ converges to 1 much faster than $\exp(\frac{1}{n})$, the correction from (1.3).

Example 3.3 (U-statistics, Example 1.3 continued). In the situation of Example 1.3, we define

$$Z = \sum_{i=\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} Z_i.$$

We first compute the central quantity in (1.3):

$$\Delta = \frac{1}{2} \binom{n}{k} \sum_{j=1}^{k-1} \binom{n-k}{j} p^{k+j} \sim C \frac{1}{k!} (np)^{k+1} \frac{1 - (np)^{k-1}}{1 - np},$$

with $C \in]0, 1[$. Thus (1.3) yields

$$\mathbb{P}(Z = 0) \leq (1 - p^k)^{\binom{n}{k}} \exp \left(C \frac{(np)^{k+1} (1 - (np)^{k-1})}{k! (1 - np) (1 - p^k)} \right) \quad (3.3)$$

for some constant C , and it is obvious that this bound is only effective, if $p = \frac{\lambda}{n}$, for some $\lambda > 0$ or even $p = o(1/n)$. This is also the situation where Theorem 1.7 gives good bounds. If $p = \lambda/n$, we choose $t \ll e^{-\lambda^k/2k!}/\Delta$ to guarantee that the second summand in (1.6) is much smaller than the first. Then, for large λ , (1.6) basically boils down to

$$\mathbb{P}(Z = 0) \leq (1 - p^k)^{\binom{n}{k}} \left(1 + \frac{p^k}{1 - p^k} \right)^{\binom{n}{k}} + o(1).$$

This is still better than the bound (3.3) obtained from (1.3), since our multiplicative error factor behaves like $e^{\lambda^k/k!}$, while the one in (3.3) is essentially $e^{\lambda^{2k}/k!}$. However, if λ is getting small, our bound gets even better (compared to (3.3)), due to the additional factor e^{-t} (with t chosen as above) in the multiplicative error term. This especially applies when $\lambda \rightarrow 0$ with n going to infinity.

In the previous examples we always estimated the sum of the covariances occurring in (1.6) by the Δ from Janson's inequality: One may wonder whether we haven't been giving away too much by such a bound. The answer is no, since in these cases the covariances of the variables are dominated by the expectations of the product. However, this need not be the case in general and it will turn out in such situations our bound is significantly better than Janson's inequality.

Example 3.4 (Random hypergraphs, Example 1.4 continued). Here we will see, that our inequality is able to outperform Janson's inequality in certain situations. To this end, in the situation of Example 1.4 let \mathfrak{Y} be the event that all edges of K_N are covered by one of the randomly chosen K_k 's. With the notation of Example 1.4

$$\mathfrak{Y} = \left\{ \sum_{1 \leq i < j \leq N} Y_{i,j} = \binom{N}{2} \right\} = \left\{ Z := \sum_{1 \leq i < j \leq N} Z_{i,j} = 0 \right\}.$$

We want to estimate the probability of \mathfrak{Y} for large n and fixed k . To this end assume we select n K_k 's as subgraphs of K_N i.i.d. at random. Then for any edge (i, j)

$$\mathbb{P}(Z_{i,j} = 1) = \left(1 - \frac{\binom{N-2}{k-2}}{\binom{N}{k}} \right)^n.$$

Indeed, $Z_{i,j} = 1$, if and only if (i, j) is not contained in any of the n independent K_k 's and this happens with probability $\left(1 - \frac{\binom{N-2}{k-2}}{\binom{N}{k}}\right)^n =: p(N, n, k) =: p$. For N large, and k fixed,

$$p \sim \left(1 - \frac{k(k-1)}{N^2}\right)^n.$$

To compute Δ from Theorem 1.5 observe that indeed none of the $Z_{i,j}$ are uncorrelated, but that their correlation is extremely weak. As a matter of fact, there are two different types of correlations between $Z_{i,j}$ and $Z_{i',j'}$ depending on whether the edges (i, j) and (i', j') share a vertex or not. Assume (i, j) and (i', j') share a vertex, e.g. $j = j'$. Then with $Y_{i,j}^r$ denoting the indicator that (i, j) is covered by the r 'th independent copy of K_k

$$\begin{aligned} \mathbb{P}(Z_{i,j}Z_{i',j} = 1) &= \mathbb{P}(Y_{i,j} + Y_{i',j} = 0) \\ &= \mathbb{P}\left(\bigcap_{r=1}^n \{Y_{i,j}^r + Y_{i',j}^r = 0\}\right) \\ &= \left(\frac{\binom{N-3}{k}}{\binom{N}{k}} + 3\frac{\binom{N-3}{k-1}}{\binom{N}{k}} + \frac{\binom{N-3}{k-2}}{\binom{N}{k}}\right)^n, \end{aligned}$$

the different terms coming from the five possible cases:

- choose k indices in $\{1, \dots, N\} \setminus \{i, i', j\}$,
- choose i and then $k-1$ indices in $\{1, \dots, N\} \setminus \{i, i', j\}$,
- choose i' and then $k-1$ indices in $\{1, \dots, N\} \setminus \{i, i', j\}$,
- choose j and then $k-1$ indices in $\{1, \dots, N\} \setminus \{i, i', j\}$,
- choose i and i' and then $k-2$ indices in $\{1, \dots, N\} \setminus \{i, i', j\}$.

In the same way, if (i, j) and (i', j') do not share a vertex, we have

$$\begin{aligned} \mathbb{P}(Z_{i,j}Z_{i',j'} = 1) &= \mathbb{P}\left(\bigcap_{r=1}^n \{Y_{i,j}^r + Y_{i',j'}^r = 0\}\right) \\ &= \left(\frac{\binom{N-4}{k}}{\binom{N}{k}} + 4\frac{\binom{N-4}{k-1}}{\binom{N}{k}} + 4\frac{\binom{N-4}{k-2}}{\binom{N}{k}}\right)^n \end{aligned}$$

Thus

$$\begin{aligned} \Delta &= \frac{1}{2} \sum_{\substack{(i,j),(i',j'), \\ (i,j) \neq (i',j')}} \mathbb{E}(Z_{i,j}Z_{i',j'}) \\ &= \binom{N}{2} \left((N-2) \left(\frac{\binom{N-3}{k}}{\binom{N}{k}} + 3\frac{\binom{N-3}{k-1}}{\binom{N}{k}} + \frac{\binom{N-3}{k-2}}{\binom{N}{k}} \right)^n \right. \\ &\quad \left. + \frac{1}{2} \binom{N-2}{2} \left(\frac{\binom{N-4}{k}}{\binom{N}{k}} + 4\frac{\binom{N-4}{k-1}}{\binom{N}{k}} + 4\frac{\binom{N-4}{k-2}}{\binom{N}{k}} \right)^n \right) \end{aligned}$$

Now suppose $n = \lambda N^2$ for some $\lambda > 0$ and that N is large. Observe that for n smaller than this, we cannot hope for any kind of Poisson approximation for the probability in question, since we simply do not have enough edges from the K_k 's to cover K_N . For n as large as $n = \lambda N^2$, or larger, the Poisson approximation becomes more likely, since then the probability not to cover an edge, becomes small.

We will now expand the exponentials in the computation of Δ . Let us start with $k = 3$, such that only the first of the two summands for Δ is present. We obtain:

$$\begin{aligned} \frac{\binom{N-3}{k}}{\binom{N}{k}} &= \left(1 - \frac{k}{N}\right)\left(1 - \frac{k+1}{N}\right)\left(1 - \frac{k+2}{N}\right)\frac{N}{N-1}\frac{N}{N-2} \\ &= 1 - 3\frac{k}{N} + \frac{1}{N^2}(3k^2 - 3k) + O\left(\frac{1}{N^3}\right) \end{aligned}$$

and

$$\begin{aligned} \frac{\binom{N-3}{k-1}}{\binom{N}{k}} &= \frac{(N-k)(N-k-1)k}{N(N-1)(N-2)} \\ &= \left(1 - \frac{k}{N}\right)\left(1 - \frac{k+1}{N}\right)\frac{k}{N}\left(1 + \frac{1}{N} + O\left(\frac{1}{N^2}\right)\right)\left(1 + \frac{2}{N} + O\left(\frac{1}{N^2}\right)\right) \\ &= \frac{k}{N}\left(1 - 2\frac{k-1}{N} + O\left(\frac{1}{N^2}\right)\right) \end{aligned}$$

as well as

$$\frac{\binom{N-3}{k-2}}{\binom{N}{k}} = \frac{(N-3)!}{N!} \frac{(N-k)!}{(N-k-1)!} \frac{k!}{(k-2)!} = \frac{(N-k)k(k-1)}{N(N-1)(N-2)} = \frac{k(k-1)}{N^2} + O\left(\frac{1}{N^3}\right).$$

Thus

$$\begin{aligned} \mathbb{P}(Z_{i,j}Z_{i',j} = 1) &= \left(\frac{\binom{N-3}{k}}{\binom{N}{k}} + 3\frac{\binom{N-3}{k-1}}{\binom{N}{k}} + \frac{\binom{N-3}{k-2}}{\binom{N}{k}} \right)^n \\ &= \left(1 - \frac{2}{N^2}k(k-1) + O\left(\frac{1}{N^3}\right) \right)^n \end{aligned}$$

As for $n = \lambda N^2$ the probability p is of constant order we obtain from Janson's inequality (1.3) for $k = 3$

$$\begin{aligned} \mathbb{P}(Z = 0) &\leq (1-p)^{\binom{N}{2}} \exp \left(CN^3 \left(1 - \frac{2}{N^2}k(k-1) + O\left(\frac{1}{N^3}\right) \right)^n \right) \\ &\leq (1-p)^{\binom{N}{2}} \exp (CN^3 e^{-12\lambda}) \end{aligned}$$

for N large enough. This is, of course, a useless bound, since the right hand side is converging to infinity.

On the other hand, for $\text{Cov}(Z_{i,j}, Z_{i',j})$ we compute up to terms of order n/N^4

$$\text{Cov}(Z_{i,j}, Z_{i',j}) = \frac{n(k^3 - 3k^2 + 2k)}{N^3} + O\left(\frac{n}{N^4}\right).$$

Thus

$$\sum_{(i,j) \sim (i',j')} \text{Cov}(Z_{i,j}, Z_{i',j'}) = C\lambda N^2.$$

Obviously, the bound (1.4) from Theorem 1.6 by Boutsikas and Koutras is useless in this example, since the sum of the covariances diverges. With the factor t^2 in front of the covariances, our bound (1.6) gives for any $t > 0$.

$$\mathbb{P}(Z = 0) \leq (1-p)^{\binom{N}{2}} \left(1 + e^{-t} \frac{p}{1-p}\right)^{\binom{N}{2}} + Ct^2 \lambda N^2.$$

If we now choose t extremely small, e.g $t = e^{-N^3}$, it is evident that the second summand on the right is asymptotically negligible with respect to the first. This first summand is estimated as

$$(1-p)^{\binom{N}{2}} \left(1 + e^{-t} \frac{p}{1-p}\right)^{\binom{N}{2}} \leq (1-p)^{\binom{N}{2}} \left(1 + \frac{e^{-6\lambda}}{1 - e^{-6\lambda}}\right)^{\binom{N}{2}}$$

Thus (1.6) gives

$$\mathbb{P}(Z = 0) \leq \left((1-p) \left(1 + \frac{e^{-6\lambda}}{1 - e^{-6\lambda}}\right) \right)^{\binom{N}{2}}$$

In particular for λ large, this is not only of the right order, but also very close to the lower bound $(1-p)^{\binom{N}{2}}$.

For $k \geq 4$ one can compute that the covariances are now of order $1/N^2$ (which we spare ourselves as the computations are similar to ones above), but as a matter of fact, a similar choice of t leads to a similar result as in the case $k = 3$.

To conclude, the above examples show that for positive correlations our method gives comparable results to that of Janson's inequality. Even more is true: There are examples where our methods gives useful bounds, while (1.3) does not. These examples occur, if many of the random variables are correlated, but correlation is very weak. In this case Δ can tend to be so large, that (1.3) does not yield reasonable bounds. Moreover, when the sum of the covariances is too large, the last example illustrates also the situation where our bound gives an interesting approximation of the independent case, while the bound by Boutsikas and Koutras doesn't apply.

REFERENCES

- [1] B. K. Aliabadi, C. Berrou, V. Gripon, and X. Jiang. Learning sparse messages in networks of neural cliques. *IEEE Transactions on Neural Networks and Learning Systems*, August 2012. To appear.
- [2] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.
- [3] M. Behrisch, A. Coja-Oghlan, and M. Kang. The order of the giant component of random hypergraphs. *Random Structures Algorithms*, 36(2):149–184, 2010.
- [4] M. Behrisch, A. Coja-Oghlan, and M. Kang. Local limit theorems for the giant component of random hypergraphs. *Combin. Probab. Comput.*, 23(3):331–366, 2014.
- [5] R. Boppona and J. Spencer. A useful elementary correlation inequality. *J. Combin. Theory Ser. A*, 50(2):305–307, 1989.
- [6] M. V. Boutsikas and M. V. Koutras. A bound for the distribution of the sum of discrete associated or negatively associated random variables. *Ann. Appl. Probab.*, 10(4):1137–1150, 2000.
- [7] A. Coja-Oghlan, C. Moore, and V. Sanwalani. Counting connected graphs and hypergraphs via the probabilistic method. *Random Structures Algorithms*, 31(3):288–329, 2007.
- [8] J. T. Cox and G. Grimmett. Central limit theorems for associated random variables and the percolation model. *Ann. Probab.*, 12(2):514–528, 1984.
- [9] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.
- [10] J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *Ann. Math. Statist.*, 38:1466–1474, 1967.
- [11] V. Gripon and C. Berrou. Sparse neural networks with large learning diversity. *IEEE Transactions on Neural Networks*, 22(7):1087–1096, July 2011.
- [12] S. Janson. Poisson approximation for large deviations. *Random Structures Algorithms*, 1(2):221–229, 1990.
- [13] S. Janson. New versions of Suen’s correlation inequality. In *Proceedings of the Eighth International Conference “Random Structures and Algorithms” (Poznan, 1997)*, volume 13, pages 467–483, 1998.
- [14] J. Kahn and N. Weininger. Positive association in the fractional fuzzy Potts model. *Ann. Probab.*, 35(6):2038–2043, 2007.
- [15] N. Karasik. *Results on the Sums of Associated Random Variables Motivated by the Parabolic Anderson Model*. ProQuest LLC, Ann Arbor, MI, 2013. Thesis (Ph.D.)—University of California, Irvine.
- [16] M. Karoński and T. Łuczak. The phase transition in a random hypergraph. *J. Comput. Appl. Math.*, 142(1):125–135, 2002. Probabilistic methods in combinatorics and combinatorial optimization.
- [17] C. M. Newman. Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.*, 74(2):119–128, 1980.
- [18] C. M. Newman and A. L. Wright. An invariance principle for certain dependent sequences. *Ann. Probab.*, 9(4):671–675, 1981.
- [19] C. M. Newman and A. L. Wright. Associated random variables and martingale inequalities. *Z. Wahrsch. Verw. Gebiete*, 59(3):361–371, 1982.
- [20] P. E. Oliveira. An exponential inequality for associated variables. *Statist. Probab. Lett.*, 73(2):189–197, 2005.
- [21] P. E. Oliveira. *Asymptotics for associated random variables*. Springer, Heidelberg, 2012.

- [22] B. L. S. P. Rao and I. Dewan. A general method of density estimation for associated random variables. *Nonparametric Statistics*, 10:405–420, 1999.
- [23] D. Romik. Connectivity Patterns in Loop Percolation I: the Rationality Phenomenon and Constant Term Identities. *Comm. Math. Phys.*, 330(2):499–538, 2014.
- [24] W.-C. S. Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures Algorithms*, 1(2):231–242, 1990.
- [25] S. H. Sung. A note on the exponential inequality for associated random variables. *Statist. Probab. Lett.*, 77(18):1730–1736, 2007.
- [26] G. Xing and S. Yang. Notes on the exponential inequalities for strictly stationary and positively associated random variables. *J. Statist. Plann. Inference*, 138(12):4132–4140, 2008.
- [27] G. Xing and S. Yang. Some exponential inequalities for positively associated random variables and rates of convergence of the strong law of large numbers. *J. Theoret. Probab.*, 23(1):169–192, 2010.
- [28] S.-c. Yang and M. Chen. Exponential inequalities for associated random variables and strong laws of large numbers. *Sci. China Ser. A*, 50(5):705–714, 2007.

(Matthias Löwe) FACHBEREICH MATHEMATIK UND INFORMATIK, UNIVERSITY OF MÜNSTER,
EINSTEINSTRASSE 62, 48149 MÜNSTER, GERMANY

E-mail address, Matthias Löwe: `maloewe@math.uni-muenster.de`

(Franck Vermet) LABORATOIRE DE MATHÉMATIQUES, UMR CNRS 6205, UNIVERSITÉ DE
BRETAGNE OCCIDENTALE, 6, AVENUE VICTOR LE GORGEU, CS 93837, F-29238 BREST
CEDEX 3, FRANCE

E-mail address, Franck Vermet: `Franck.Vermet@univ-brest.fr`